We claim:

1. In a computer data processing system, a method for clustering data in a database comprising:

a) providing a database having a number of data records having both discrete and continuous attributes;

b) grouping together data records from the database which have specified discrete attribute configurations;

c) clustering data records having the same or similar specified discrete attribute configuration based on the continuous attributes to produce an intermediate set of data clusters; and

d) merging together clusters from the intermediate set of data clusters to produce a clustering model.

2. The method of claim 1 wherein the clustering model includes a table of probabilities for the discrete data attributes of the data records for a cluster and wherein the cluster model for continuous data attributes comprises a mean and a covariance for each cluster.

3 . The method of claim 1 wherein the process of merging of intermediate clusters is ended when a specified number of clusters has been formed.

4. The method of claim 1 wherein the step of merging of intermediate clusters is ended when a distance between intermediate clusters is greater than a specified minimum distance.

5. The method of claim 1 wherein the discrete attributes are Boolean and similarity between configurations is based on a distance between bit patterns of the discrete attributes.

6. The method of claim 1 wherein one or more of the discrete attributes have more than two possible values and comprising the step of subdividing a discrete attribute having more than two possible values into multiple Boolean value attributes.

7. The method of claim 5 wherein the step of identifying configurations includes tabulating data records having the same discrete attribute bit pattern and combining the data records from similar configurations before clustering the data records so tabulated based on the continuous attributes.

8. In a computer data processing system, a method for clustering data in a database comprising:

a) providing a database having a number of data records having both discrete and continuous attributes;

b) counting data records from the database which have the same discrete attribute configuration and identifying a first set of configurations wherein the number of data records of each configuration of said first set of configurations exceeds a threshold number of data records;

c) adding data records from the database not belonging to one of the first set of configurations with a configuration within said first set of configurations to produce a subset of records from the database belonging to configurations in the first set of configurations; and

d) clustering the subset of records contained within at least some of the first set of configurations based on the continuous data attributes of records contained within that first set of configurations to produce a clustering model.

9. The method of claim 8 wherein the clustering model includes a table of probabilities for the discrete data attributes of the data records for a cluster and wherein the cluster model for continuous data attributes comprises a mean and a covariance for each cluster.

10 . The method of claim 8 wherein an added record not contained within the first set of configurations is added to one of said first set of configurations based on a distance between a smaller configuration to which said added record belongs during counting of records in different configurations.

11. The method of claim 8 wherein the clustering of records from a configuration based on continuous data attributes results in a variable number of clusters for each configuration based on the number of records in said configuration.

12. The method of claim 8 wherein the clustering of records from records falling within a configuration of the first set results in a number of intermediate clusters which are merged together to form the cluster model.

13. The method of claim 12 wherein intermediate clusters are merged together based on a distance between clusters that is determined based on both continuous and discrete attributes of said intermediate clusters.

14. The method of claim 13 wherein the merging of intermediate clusters is performed until a specified number of clusters are contained in the cluster model.

15. The method of claim 13 wherein the merging of intermediate clusters is performed until a distance between two closest clusters is greater than a threshold distance.

16. The method of claim 8 wherein a list of records of each configuration in the first set of configurations is maintained as data records are accessed from the database.

17. The method of claim 8 where the clustering based on the continuous attributes of records within a configuration is performed using expectation maximization clustering of the continuous attributes.

18. A data processing system comprising:

a) a storage medium for storing a database having a number of data records having both discrete and continuous attributes;

b) a computer for evaluating data records from the database and building a clustering model that describes data in the database; and

c) a database management system including a component for selectively retrieving data records from the database for evaluation by the computer;

d) said computer including a stored program for i) grouping together data records from the database which have specified discrete attribute configurations; ii) clustering data records having the same or a similar specified discrete attribute configuration based on the continuous attributes to produce an intermediate set of data clusters; and iii)

merging together clusters from the intermediate set of data clusters to produce a clustering model.

19. The system of claim 18 wherein the computer includes a rapid access storage for maintaining a list of data records from the database for data records having a specified discrete attribute configuration to facilitate clustering of the data records based on their continuous attributes.

20. The data processing system of claim 18 wherein the database management system comprises means for subdividing discrete attributes having more than two possible values into multiple Boolean value attributes having two possible values.

21. The system of claim 18 wherein the rapid access storage of said computer includes a data structure for storing a clustering model.

22. A computer readable medium containing stored instructions for clustering data in a database comprising instructions for :

a) reading records from a database having a number of data records having both discrete and continuous attributes;

b) grouping together data records from the database which have specified discrete attribute configurations;

c) clustering data records having the same or similar specified discrete attribute configuration based on the continuous attributes to produce an intermediate set of data clusters; and

d) merging together clusters from the intermediate set of data clusters to produce a clustering model.

23. The computer readable medium of claim 22 including instructions for maintaining a clustering model that includes a table of probabilities for the discrete data attributes of the data records for a cluster and wherein the cluster model for continuous data attributes comprises a mean and a covariance for each cluster.

24 . The computer readable medium of claim 22 wherein the instructions end the process of merging of intermediate clusters when a specified number of clusters has been formed.

25. The computer readable medium of claim 22 wherein the instructions end the process of merging intermediate clusters when a distance between intermediate clusters is greater than a specified minimum distance.

26. The computer readable medium of claim 22 wherein the discrete attributes are Boolean and the instructions determine similarity between configurations based on a distance between bit patterns of the discrete attributes.

27. The computer readable medium of claim 22 wherein the instructions identify configurations by tabulating data records having the same discrete attribute bit pattern and combining the data records from similar configurations before clustering the data records so tabulated based on the continuous attributes.

28. The computer readable medium of claim 22 wherein the clustering of records from a configuration based on continuous data attributes produces a variable number

of the intermediate clusters for each configuration based on the number of records in said configuration.

29. The computer readable medium of claim 22 wherein the instructions maintain a list of records of each configuration as data records are accessed from the database.

30. The computer readable medium of claim 22 wherein the instructions cluster records within a configuration based on the continuous attributes of records within that configuration using expectation maximization clustering of the continuous attributes.

31. The computer readable medium of claim 30 where records are assigned to a single cluster during the expectation maximization clustering process.